MAKING BINARY DECISIONS BASED ON THE POSTERIOR PROBABILITY DISTRIBUTION ASSOCIATED WITH TOMOGRAPHIC RECONSTRUCTIONS

Kenneth M. Hanson
Los Alamos National Laboratory
MS P940
Los Alamos, New Mexico 87545 USA

ABSTRACT. An optimal solution to the problem of making binary decisions about a local region of a reconstruction is provided by the Bayesian method. The decision is made on the basis of the ratio of the posterior probabilities for the two hypotheses. The full Bayesian procedure requires an integration of the posterior probability over all possible values of the image outside the local region being analyzed. In the present work, this full treatment is replaced by the maximum value of the posterior probability obtained when the exterior region is varied, but the interior is fixed at both hypothesized functional forms. A Monte Carlo procedure is employed to evaluate the usefulness of the technique in a signal-known-exactly detection task in a noisy four-view tomographic reconstruction situation.

1. Introduction

When interpreting reconstructed images, it is often desired to make a decision about a small region of interest without regard to the rest of the image. A standard approach to this problem might be to reconstruct the full image from the available data and then make the decision on the basis of how closely the reconstruction resembled the suspected object in the region of interest. Such an approach is not guaranteed to yield an optimal decision.

We desire a computational method that achieves the optimal performance of a binary decision task. Such an 'ideal observer' has been useful in the past to help define the ultimate precision with which one can interpret data of a given type (Hanson, 1980, 1983; Wagner *et al.*, 1989; Burgess *et al.*, 1984a, 1984b; Burgess, 1985). A fully Bayesian approach is proposed in which the decision is based on the posterior probability. Much of this work is based on the Bayesian concepts developed by Gull and Skilling and their colleagues (Gull, 1989a, 1989b; Gull and Skilling, 1989; Skilling, 1989), albeit under the assumption of a Gaussian distribution for the prior probability rather than their preferred entropic form.

Examples of this Bayesian decision procedure are presented for a computed tomographic situation in which a nonnegativity constraint on the image is incorporated. The performance of the comprehensive Bayesian procedure is compared to that of the traditional two-step approach using a Monte Carlo simulation of the entire imaging process, including the decision process (Hanson, 1988a, 1990a). The present work is an extension of the previous study by Hanson (1991).

2. The Bayesian Approach

We briefly present the concepts of the Bayesian approach relevant to our problem. Much more complete discussions of the fundamentals can be found in other contributions to the proceedings of this workshop series. The essence of the Bayesian approach is the posterior probability, which is assumed to summarize the full state of knowledge concerning a given situation. The posterior probability properly combines the likelihood, which is based on the recently acquired measurements, with the prior probablity, which subsumes all information existing before the new data are acquired.

Making a binary decision is the simplest possible type of hypothesis testing, because there are just two alternative models between which to choose. According to basic probability theory, binary decisions should be made on the basis of the ratio of the probabilities for each of the hypotheses (Van Trees, 1968; Whalen, 1971). In the context of Bayesian analysis, then, a binary decision should be based on the ratio of posterior probabilities. This decision strategy is altered when there are asymmetric cost functions, indicating a difference in the relative value of making correct versus incorrect decisions for each state of truth.

When a continuum of possible outcomes exists, as in the estimation of one (or many) continuous parameters, the best possible choice of parameter values depends upon the type of cost function that is appropriate. It may be argued that for general analyses, the most appropriate rule is to find the parameters that maximize the posterior probability, which is called the maximum *a posteriori* (MAP) solution (Van Trees, 1968).

In many problems there exist parameters that may be necessary to fully describe the solution, but whose values are of no interest. These unnecessary parameters can transform a simple hypothesis test into one of testing composite hypotheses. In such cases the proper approach is to integrate the probability density distribution over these unwanted variables. The result of this integration is called the marginal probability.

2.1 Posterior Probability

We assume that there exists a scene that can be adequately represented by an orderly array of N pixels. We are given M discrete measurements that are linearly related to the original image amplitudes. These measurements are assumed to be degraded by additive noise with a known covariance matrix $\mathbf{R_n}$, which describes the correlations between noise fluctuations. The measurements, represented by a vector of length M, can be written as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{1}$$

where $\mathbf{f}$ is the original image vector of length N, $\mathbf{n}$ is the random noise vector, and $\mathbf{H}$ is the measurement matrix. In computed tomography the jth row of $\mathbf{H}$ describes the weight of the contribution of image pixels to the jth projection measurement.

Because the probability is a function of continuous parameters, namely the N pixel values of the image and the M data values, it is actually a probability density, designated by a small $p()$. The negative logarithm of the posterior probability is given by

$$-\log\left[p(\mathbf{f}|\mathbf{g})\right] = \phi(\mathbf{f}) = \Lambda(\mathbf{f}) + \Pi(\mathbf{f}), \tag{2}$$

where the first term comes from the likelihood and the second term from the prior probability. For additive Gaussian noise, the negative log(likelihood) is just half of chi-squared

$$-\log\left[p(\mathbf{g}|\mathbf{f})\right] = \Lambda(\mathbf{f}) = \frac{1}{2}\chi^2 = \frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})^T\mathbf{R_n}^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) , \tag{3}$$

which is quadratic in the residuals. Instead of a Gaussian distribution assumed here, the Poisson distribution is often a better model for expected measurement fluctuations. The choice should be based on the statistical characteristics of the measurement noise, which we assume are known *a priori*.

The prior-probability distribution should incorporate as much as possible the known characteristics of the original image. We use a Gaussian distribution for the prior, whose negative logarithm may be written as

$$-\log[p(\mathbf{f})] = \Pi(\mathbf{f}) = \frac{1}{2}(\mathbf{f}-\bar{\mathbf{f}})^{\mathrm{T}}\mathbf{R}_{\mathbf{f}}^{-1}(\mathbf{f}-\bar{\mathbf{f}}) \,, \tag{4}$$

where $\bar{\mathbf{f}}$ is the mean and $\mathbf{R}_{\mathbf{f}}$ is the covariance matrix of the prior-probability distribution. As we have done before (Hanson and Myers, 1990c), we invoke the prior knowledge that the image $\mathbf{f}$ cannot possess any negative components and institute nonnegativity as a separate constraint.

The Bayesian approach is not based on any particular choice of prior. Another choice for prior, ubiquitous at this workshop, is that of entropy. The entropic prior has been argued by Skilling (1989) to play a unique role for additive positive distributions. Whatever prior is used, its strength affects the amount the reconstruction is offset from the true image (Hanson, 1990b; Myers and Hanson, 1990). It is important to understand the characteristics of solutions obtained regardless of the prior chosen. It is recognized that the prior provides the regularization essential to solving ill-posed problems (Nashed, 1981; Titterington, 1985), which arise because $\mathbf{H}$ possesses a null-space (Hanson and Wecksung, 1983; Hanson, 1987).

Given the posterior probability, we have the means for deciding between two possible images $\mathbf{f}_1$ and $\mathbf{f}_2$. Under the standard assumption that the noise is stationary and uncorrelated $\mathbf{R}_{\mathbf{n}} = \mathrm{diag}(\sigma_{\mathbf{n}}^2) = (1.0)^2$, the decision should be based on the posterior probabilities, or equivalently, the difference of their logarithms

$$\psi_{21} = \phi(\mathbf{f}_2) - \phi(\mathbf{f}_1) = \frac{1}{2\sigma_{\mathbf{n}}^2}\left[2\mathbf{g}^{\mathrm{T}}(\mathbf{g}_1 - \mathbf{g}_2) + |\mathbf{g}_2|^2 - |\mathbf{g}_1|^2\right] + \text{constant} \,, \tag{5}$$

where $\mathbf{g}_k = \mathbf{H}\mathbf{f}_k$. The only part of this expression that depends on the data is the inner product between $\mathbf{g}$ and $(\mathbf{g}_1 - \mathbf{g}_2)$. We note that this inner product represents the familiar cross correlation between the data and the difference between the alternative signals, which is called the matched filter. The constant in Eq. (5) depends solely on $\mathbf{f}$, $\mathbf{f}_1$, and $\mathbf{f}_2$. It provides an offset to ψ_{21} indicating a prior preference for one of the two choices. As ψ_{21} is linearly dependent on the data, it too has a Gaussian-shaped probability distribution.

2.2 Reconstruction Problem

In the reconstruction problem, we seek to estimate all pixel values in the original scene. An appropriate Bayesian solution to this problem is the image that maximizes the posterior probability or, equivalently, minimizes the negative logarithm of the posterior probability. For the **unconstrained** MAP solution $\hat{\mathbf{f}}$, it is necessary that

$$\nabla_{\mathbf{f}}\phi = \mathbf{R}_{\mathbf{f}}^{-1}(\mathbf{f}-\bar{\mathbf{f}}) + \mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{n}}^{-1}(\mathbf{g}-\mathbf{H}\mathbf{f}) = 0 \,. \tag{6}$$

However, under the constraint that the solution should be nonnegative, the derivative with respect to f_i must be zero only when $f_i > 0$; a negative derivative is permissible on the

boundary $f_i = 0$. In computed tomography (CT), the matrix operation $\mathbf{H}^T$ is the familiar backprojection process.

A consequence of the prior is to pull the reconstruction away from the actual value in the original image, an effect studied by Hanson (1990b) in unconstrained tomographic reconstructions. The extent of bias depends on the relative weights of the two terms in Eq. (2). As the prior contribution vanishes, the MAP result approaches the maximum likelihood (or least-square residual) solution.

2.3 Analysis of a Local Region

Instead of asking for an estimate of the original image, suppose that we ask a different question: which of two objects exists at a specific location in the image? The rest of the image suddenly becomes irrelevant. To address this question, we assume that within the image domain, a local region $\mathcal{D}$ is to be analyzed. Inside $\mathcal{D}$ the image $\mathbf{f}$ is assumed to be given by either $\mathbf{f}_{\mathcal{D}1}$, and $\mathbf{f}_{\mathcal{D}2}$. Now the parameters in the problem are not the full set of image values $\mathbf{f}$, but rather $\mathbf{f}_{\mathcal{E}}$, the image values in the disjoint exterior region $\mathcal{E}$ and the two possible choices for $\mathcal{D}$. With Bayes' law the posterior probability may be written as $p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k})p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}) \propto p(\mathbf{g}|\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k})p(\mathbf{f})$. In the last step we have chosen to avoid explicit specification of a prior on $\mathbf{f}_{\mathcal{D}k}$, allowing it to be implicitly included in the general prior for $\mathbf{f}$.

As the new question regards only the region $\mathcal{D}$, the image values $\mathbf{f}_{\mathcal{E}}$ outside $\mathcal{D}$ are irrelevant. The Bayesian approach specifies that we integrate the posterior probabilities over the unwanted parameters of the problem, namely over the image values outside $\mathcal{D}$. If the problem at hand is to decide between two possible subimages, $\mathbf{f}_{\mathcal{D}1}$ or $\mathbf{f}_{\mathcal{D}2}$, the decision variable should be the ratio of the two marginal posterior probabilities (Van Trees, 1968; Whalen, 1971), or equivalently its logarithm

$$\psi = \log\left[\frac{\int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}1}|\mathbf{g})d\mathbf{f}}{\int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}2}|\mathbf{g})d\mathbf{f}}\right] , \tag{7}$$

where the integrals are to be carried out only over the external region $\mathcal{E}$ and include all possible image values not disallowed by constraints. Within the context of Bayesian analysis, this decision variable logically follows from the statement of the problem. Hence, we assert that it should yield optimal decisions. The ideal observer uses Eq. (7) to make binary decisions regarding a local region.

Under certain circumstances these integrals may be difficult to calculate accurately. However, when dealing with the Gaussian prior- and likelihood-probability density distributions presented in Sec. 2.1, we expect the posterior-probability density $p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})$ to decrease rapidly from a unique maximum. Using $\hat{\mathbf{f}}_{\mathcal{E}k}$ to designate the image in the exterior region that maximizes the posterior probability for the subimage $\mathbf{f}_{\mathcal{D}k}$, we are prompted to rewrite the above ratio as,

$$\psi = \log\left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}1}, \mathbf{f}_{\mathcal{D}1}|\mathbf{g})\, K_1}{p(\hat{\mathbf{f}}_{\mathcal{E}2}, \mathbf{f}_{\mathcal{D}2}|\mathbf{g})\, K_2}\right] , \tag{8}$$

where the phase-space factor is

$$K_k = \frac{1}{p(\hat{\mathbf{f}}_{\mathcal{E}k}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})} \int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})d\mathbf{f} , \tag{9}$$

which accounts for the extent of the spread in f-space of the posterior-probability density distribution about its constrained peak value $p(\hat{\mathbf{f}}_{\mathcal{E}k}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})$.

Generally $\hat{\mathbf{f}}_{\mathcal{E}1} \neq \hat{\mathbf{f}}_{\mathcal{E}2}$, because a change in model parameters describing the interior region alters the projections, implying that a different exterior image will minimize the posterior probability. In many situations, however, replacing the local region of the MAP solution with either $\mathbf{f}_{\mathcal{D}2}$ or $\mathbf{f}_{\mathcal{D}1}$ may have little effect on the predicted projection values. Then, $p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})$ is independent of $\mathbf{f}_{\mathcal{D}k}$ and, to good approximation, $\hat{\mathbf{f}}_{\mathcal{E}1} = \hat{\mathbf{f}}_{\mathcal{E}2} = \hat{\mathbf{f}}_{\mathcal{E}}$, so both K factors in Eq. (8) are the same and

$$\psi = \log\left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}1}|\mathbf{g})}{p(\hat{\mathbf{f}}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}2}|\mathbf{g})}\right] . \tag{10}$$

In these situations, the decision variable can be given adequately by the change in the log(posterior probability) induced by replacing the MAP solution $\hat{\mathbf{f}}$ in $\mathcal{D}$ with the two models, leaving the exterior region unchanged.

For unconstrained solutions of Eq. (6), the K factor is independent of $\mathbf{f}_{\mathcal{D}k}$, because the shape of the Gaussian posterior-probability distribution is governed by the full curvature of ϕ, namely $\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{R}_{\mathbf{f}}^{-1}$. Then the K factors in Eq. (8) cancel and

$$\psi = \log\left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}1}, \mathbf{f}_{\mathcal{D}1}|\mathbf{g})}{p(\hat{\mathbf{f}}_{\mathcal{E}2}, \mathbf{f}_{\mathcal{D}2}|\mathbf{g})}\right] . \tag{11}$$

The argument of the logarithm is called the generalized posterior-probability ratio. Equation (11) may not be a good approximation to (7) for constrained solutions, as the contribution to the phase-space K factor from the integral over each f_i depends on the relation of the peak in $\hat{\mathbf{f}}_{\mathcal{E}}$ to the constraint boundary. Nonetheless, because of its simplicity, we use Eq. (11) and reserve for the future an investigation of a better approximation.

To evaluate Eq. (11) for subimages $\mathbf{f}_{\mathcal{D}1}$ and $\mathbf{f}_{\mathcal{D}2}$, it is necessary to find the pair of exterior images, $\mathbf{f}_{\mathcal{E}1}$ and $\mathbf{f}_{\mathcal{E}2}$, that maximize the posterior-probability density. In other words, one must find the maximum *a posteriori* or MAP reconstruction in the exterior region with the image inside the local region fixed by the parameter values. To extend the binary decision problem to one in which the model parameters are to be estimated, it becomes necessary to simultaneously estimate the parameters and reconstruct the exterior region with the aim of minimizing the posterior probability.

We employ the iterative method described by Butler, Reeds, and Dawson (1981) to find the constrained MAP solutions. See (Hanson and Myers, 1991; Hanson, 1991) for more details.

3. Methodology

We demonstrate the use of the Bayesian approach to making decisions about a local region in a reconstructed image with a very simple example: detection of disks based on a very limited number of noisy projections. This binary discrimination task is employed because it is theoretically tractable, it is easy to perform the required decision-making procedure, and it is possible to summarize the results simply.

3.1 Monte Carlo Method to Evaluate Task Performance

The overall method for evaluating a reconstruction algorithm used here has been described before (Hanson, 1988a, 1990a). In this method a task performance index for a specified imaging situation is numerically evaluated. The technique is based on a Monte Carlo simulation of the entire imaging process including random scene generation, data taking, reconstruction, and performance of the specified task. The accuracy of the task performance is determined by comparison of the results with the known original scene using an appropriate figure of merit. Repetition of this process for many randomly generated scenes provides a statistically significant estimate of the performance index (Hanson, 1990a).

3.2 Specifications of Detection Tests

The imaging situation is chosen in an attempt to maximize the possible effect of reestimation of the exterior region implied by the full Bayesian treatment. The original scenes contain either one or two disks, all with amplitude 0.1 and diameter 8 pixels. The disks are randomly placed, but not overlapping, within the circle of reconstruction of diameter 64 pixels. The background level is zero. Enough scenes are generated in the testing sequence to provide 100 disks with amplitude 0.1 and 100 null disks placed in the background region.

The measurements consist of four parallel projections, each containing 64 samples, taken at 45° increments in view angle. Measurement noise is simulated by adding to each measurement a pseudorandom number taken from a Gaussian distribution with a standard deviation of 2. The peak projection value of each disk is 0.80. The signal-to-noise ratio (SNR) for the signal-known-exactly (SKE)/ background-known-exactly (BKE) detection of a disk may be easily calculated as $\left[\sum \mathrm{SNR}_i^2\right]^{1/2}$, where SNR_i is the SNR of the ith measurement, summed over all measurements that subtend the disk, yielding $\mathrm{SNR}_{\mathrm{detect}} = d' = 1.89$. To avoid aliasing artifacts in the reconstruction, the projection data used for reconstruction are presmoothed using a triangular convolution kernel with a FWHM of 3 sample spacings. As a result, the expected rms noise value in the smoothed data is reduced very nearly to 1.0. Thus for all cases studied we use the noise covariance matrix $\mathbf{R}_n = \mathrm{diag}(\sigma_n^2) = (1.0)^2$. With this assumption we are ignoring the correlations in the data caused by presmoothing.

For the Gaussian prior probability distribution we employ the ensemble mean $\bar{f}_i = 0.0031 = \text{constant}$, which is the average value of the scenes containing two disks. We assume the ensemble covariance matrix is diagonal with $\mathbf{R}_f = \mathrm{diag}(\sigma_f^2)$ and explore the effect of choosing different values of σ_f.

The stated task is to detect the presence of the disks under the assumption that the signal is known exactly (SKE) and the background is known exactly (BKE) in the 2D local region. The various strategies for making this binary decision are presented in the next section. A useful measure to summarize the performance of binary decisions is the detection index d_A, which is based on the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve is obtained in the usual way (Hanson, 1990a) from the histograms in the decision variable for the signal-known-present and the signal-known-absent tests. Once the ROC curve is generated and its area A determined, then d_A is found using $d_A = 2\ erfc^{-1}\{2(1 - A)\}$, where $erfc^{-1}$ is the inverse complement of the error function. There are good reasons for not using the detectability index d', which is based on the first and second moments of the histograms of the decision variable (Wagner *et al.*, 1990). For

a fixed number of binary tests, the relative statistical error in d_A is smallest when d_A is about 2.2 (Hanson, 1990a). The imaging situation should be arranged to keep d_A roughly between 1 and 3.5 to optimize the statistical value of the testing procedure.

3.3 Decision Strategies

For the simple binary discrimination tests performed here, only two parameters are needed to describe the model for the local region – the background level and the disk amplitude relative to the background. The background is assumed to be constant. The position and diameter of the disk are assumed to be known. The edge of the disk is linearly ramped over 2 pixels in radius to roughly match the blur caused by the reconstruction process. The local region of analysis is assumed to be circular with a diameter of 14 pixels and centered on the test position. When the disk is assumed present, the amplitude is set to 0.1 and when assumed absent, 0. The background level is 0 for both tests. Because of this choice for the model, it should be understood that all references to the amplitude of the disk implicitly mean relative to the surrounding background. In all the decision strategies, a decision variable is evaluated for each of the two hypotheses, and the difference between the two values is used to make the decision whether a disk is present or not.

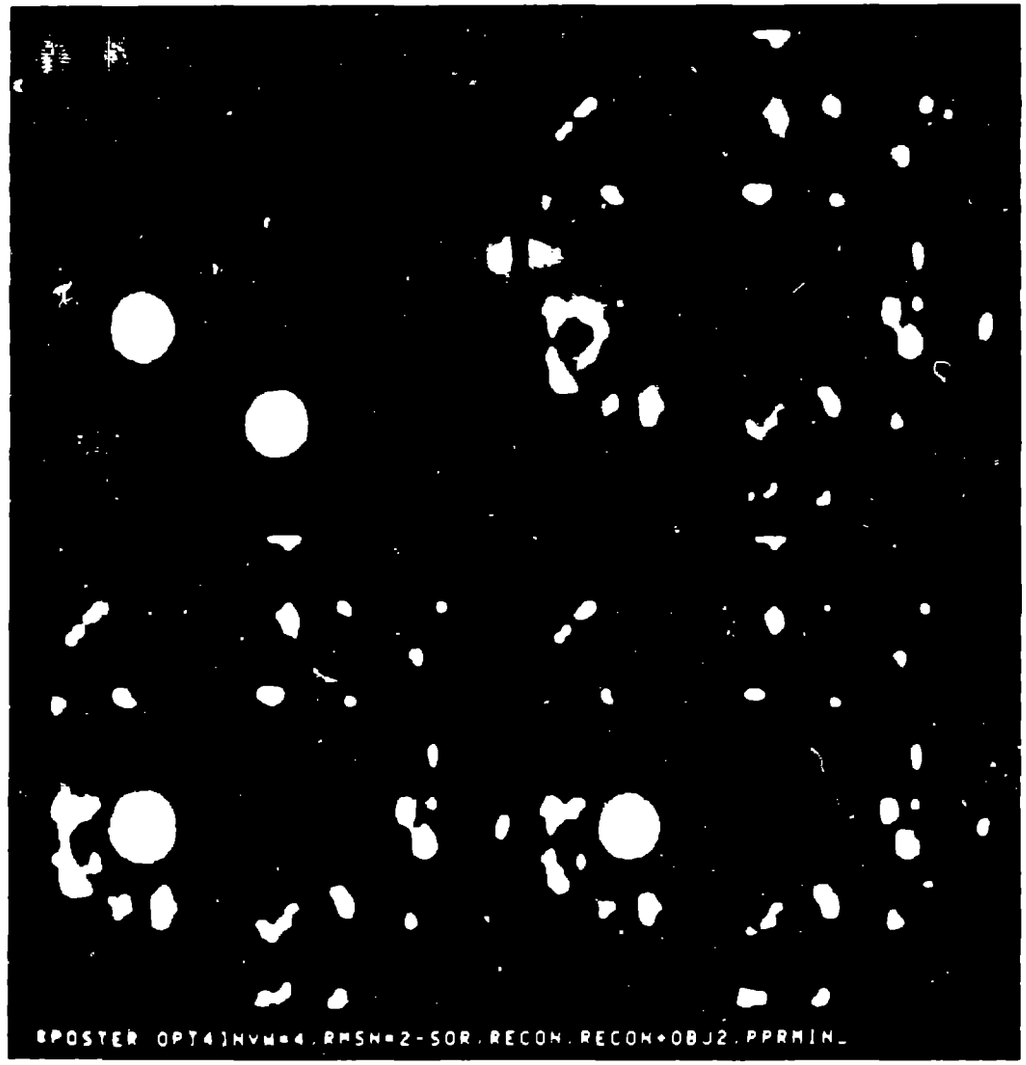
The following decision strategies are employed in this study:

Method A) In the simplest possible approach, one uses the projection data directly. The decision is based on the difference in χ^2 for the two hypotheses. Explicitly, Eq. (11) is evaluated under both hypothesized subimages for the local region of analysis $\mathcal{D}$. The image values outside the analysis region are implicitly assumed to be zero. If the background is truly zero and only one disk is present in the scene, this decision variable operates at the statistical limit attainable in the absence of prior information. However, it is obviously deficient for complex scenes as it ignores the contributions to the projections arising from features outside the local region.

Method B) By Bayesian reckoning, the best possible decision variable for local analysis is given by Eq. (7). For this method we use the approximation given by the generalized posterior-probability ratio Eq. (11), which implies that for each choice of image for D, the exterior region is reconstructed to maximize $p(\mathbf{f}_{\mathcal{E}}, \mathbf{f}_{\mathcal{D}k}|\mathbf{g})$. In actual practice, this second reconstruction step follows a preliminary constrained MAP reconstruction of the whole image as pictorially described in Fig. 1.

Method C) This method uses Eq. (10) for the decision variable based on the posterior-probability distribution associated with the MAP reconstruction. Readjustment of the reconstruction external to the analysis region for each test hypothesis is not required. This method was introduced by Gull and Skilling (1989) and studied by Myers and Hanson (1990) for an entropy prior.

Method D) Method D proceeds from the constrained MAP reconstruction $\hat{\mathbf{f}}_{\mathbf{MAP}}$ from the data. The decision variable is taken as the difference in $|\mathbf{f} - \hat{\mathbf{f}}_{\mathbf{MAP}}|^2$ for the two models hypothesized for the local region. This method was used by Hanson and Myers (1991a)to compare performance of the Rayleigh task using MAP reconstructions based on Gaussian and entropy priors. It corresponds to using a likelihood approach based on the reconstruction in which the noise fluctuations in the reconstruction are assumed to be uncorrelated and Gaussian distributed. This method therefore ignores the correlations in the posterior-probability distribution, shown in Fig. 2, that are incorporated to various degrees by methods B and C.



← - can remove bottom legend

Fig. 1. This composite image shows the process used to make the binary decision regarding the presence of a disk. The original scene (upper left) is reconstructed from four projections using constrained maximum *a posteriori* reconstruction (upper right) with ensemble standard deviation $\sigma_f = 1$. To test the possible presence of a disk, that disk is placed into the reconstruction (lower left). Then the image outside the local region of the disk is 're-reconstructed' to obtain the image (lower right) that maximizes the posterior probability with the disk present. This procedure is repeated with the same region replaced by the background value (zero). The difference in the logarithms of the two resulting posterior probabilities is used as the decision variable.

Method E) Method E also proceeds from the constrained MAP reconstruction $\hat{\mathbf{f}}_{MAP}$. Unlike the preceding methods, the amplitude and background are varied to find the combination of values that minimizes $|\mathbf{f} - \hat{\mathbf{f}}_{MAP}|^2$. In this fitting process, both the relative amplitude and the background are constrained to be nonnegative. The amplitude so deter-

Remove top title & ~~legend~~.

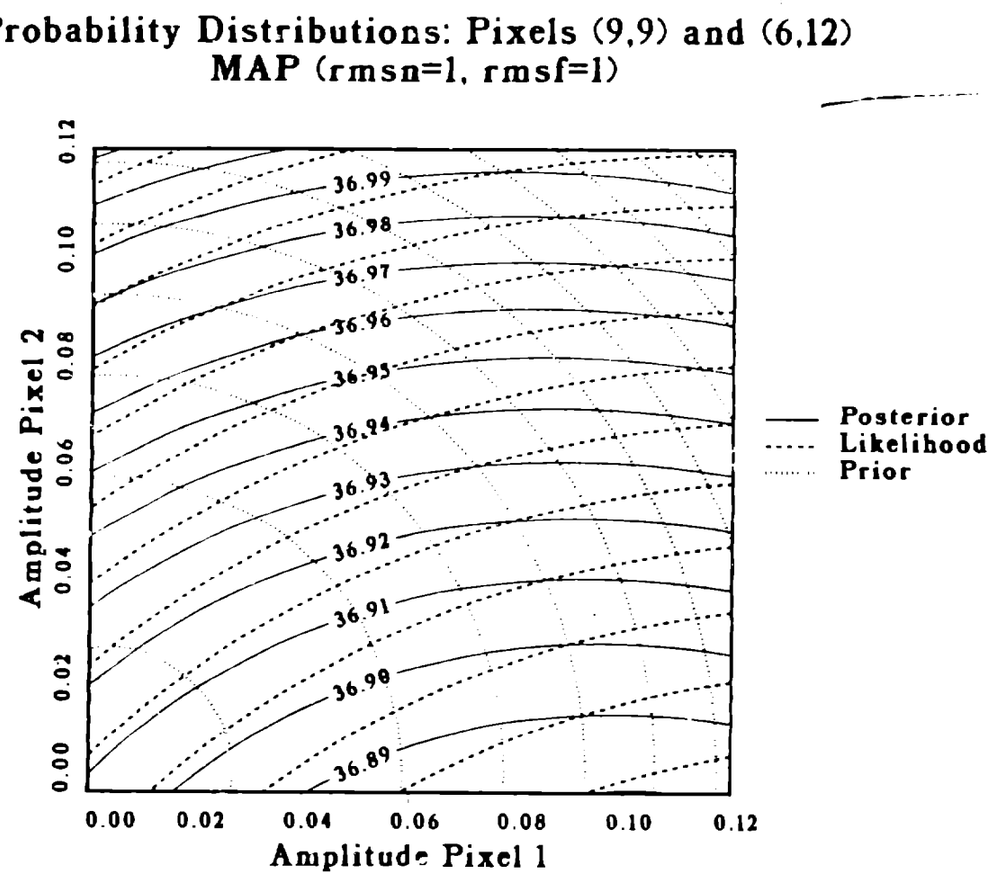


Fig. 2. Contour plot showing the correlation in −log(posterior probability) (solid line) for fluctuations in two pixel values about the MAP solution for an assumed value of the ensemble standard deviation $\sigma_f = 1.0$. The first pixel is centered on the lower middle disk in the first scene (Fig. 1) and the other is three pixels down and three pixels to the left of the first. The dotted contours are for the likelihood and the dashed contours for the prior.

mined is used as the decision variable. This method was used by Hanson in many earlier studies (1988a, 1988b, 1990a, 1990b, 1990c). It is closely related to the non-prewhitening matched filter, which would be optimal if the fluctuations in the reconstruction were uncorrelated and Gaussian distributed.

4. Example

A constrained MAP reconstruction of the first scene of the testing sequence for two disks is shown in Fig. 1. Because of the noise in the projection data, the presence of the disks in the original scene is obscured in the reconstruction. An interesting aspect of the posterior-probability approach is that one may calculate the probability of a disk being present at any location in the reconstruction. Even though the reconstruction might be zero (the lower limit decreed by the constraint of nonnegativity) throughout a certain region, the probability of a disk being present in that region is finite and calculable. By contrast, any analysis method based solely on the reconstruction would not be able to distinguish two different regions that are completely zero. This point is emphasized by the contour plot

Method	Decision Variable	d_A $\sigma_f = 0.02$	$\sigma_f = 0.1$	$\sigma_f = 0.2$	$\sigma_f = 1$
A	$\Delta\chi^2$ (use data only)	1.75	same	same	same
B	Δlog(posterior probability) (exterior re-estimated)	1.80	1.87	1.82	1.74
C	Δlog(posterior probability) (exterior fixed at $\hat{f}_{MAP}$)	1.81	1.87	1.81	1.70
D	$\Delta\|\mathbf{f} - \hat{\mathbf{f}}_{MAP}\|^2$ (use reconstruction only)	1.80	1.76	1.67	1.47
E	Disk amplitude (constrained fit to $\|\mathbf{f} - \hat{\mathbf{f}}_{MAP}\|^2$)	1.01	1.09	1.01	0.96
	RMS residual	0.914	0.823	0.774	0.725
	$\langle$amplitude$\rangle_{disk}$	0.0005	0.0068	0.0134	0.0247
	$\langle$amplitude$\rangle_{bkg}$	0.0002	0.0014	0.0024	0.0039

Table 1. Summary of the performance of the detection task for scenes containing two disks each obtained using the decision methods described in the text.

in Fig. 2, when the values of two nearby pixels are varied. The MAP solution for one of the pixels is zero, although both pixels actually fall within a disk in the original scene and should have a value 0.1. The plot shows how the prior probability shifts the posterior away from the likelihood.

The test sequences generated to demonstrate the use of posterior probability in decision making are analyzed for several different values of the ensemble covariance matrix σ_f. As we have found before (Hanson, 1989b, 1990b; Hanson and Myers, 1991a; Myers and Hanson, 1990), the performance of vision-like tasks usually varies with the parameters that control the rms residual achieved by the reconstruction algorithm. For the present MAP algorithm, that parameter is the ratio σ_f/σ_n. Recall that σ_n is fixed at its expected value of 1.0. The strength of the prior is proportional to $1/\sigma_f^2$. As the prior becomes stronger, the rms residuals of the constrained MAP reconstructions increase. The disk amplitudes, measured as the average value over each disk relative to the average over its surrounding annulus (essentially method E), are steadily reduced. These amplitudes never come close to the actual value of 0.10, probably because there are so few views, giving rise to a gigantic null space (Hanson, 1987), together with so much noise. When a Gaussian prior with $\bar{f}_i = 0$ is employed, which is nearly the case here, the MAP algorithm amounts to using minimum-norm regularization. Therefore, control of the noise, which dominates the reconstructed field, can only be achieved by reducing the sensitivity of the reconstruction (Hanson, 1990b).

Table 1 summarizes the detectability results obtained in the tests described above for the two disks/scene test. The absolute statistical accuracy of these d_A values is about 0.25. Much better accuracy should prevail in comparisons between entries in the table, however, because they are obtained by analyzing the exact same data sequence. The d_A value for the two-disk scenes based on using just the measurement data (Method A) is 1.75, in good agreement with the value of 1.89 estimated in Sec. 3.2. As only the likelihood is involved, this value is independent of σ_f. Both methods of using the posterior probability (methods B and C) provide nearly the same detectability over a large range of σ_f values. Perhaps

this consistent behavior stems from the ability of the posterior probability to fully retain the available information even though σ_f changes. There seems to be little advantage to re-estimation of the exterior of the local region to minimize the posterior probability implied by Eq. (11) in this imaging situation. There is a slight trend toward better detectability as σ_f gets smaller. The force of regularization imposed by the prior is overwhelming at $\sigma_f = .02$. For example, the reconstruction values lie between 0.0005 and 0.0046; the nonnegativity constraint is not even engaged. We observe very similar trends for the single disk scenes as well.

The methods based on the posterior probability yield sightly better ($\approx 10\%$) detectabilities than method D, which is based only on the reconstruction under the assumption that the noise uncertainty is uncorrelated and stationary. Basing the decision on the estimated disk amplitude (method E) significantly ($\approx 45\%$) reduces detectability compared to the other methods.

For **unconstrained** MAP with $\sigma_f = 0.1$, the d_A values are nearly the same as those in Table 1, so the nonnegativity constraint has little effect on detectability in the present situation. In previous work involving a limited number of views, we have seen remarkable improvements in detectability wrought by the nonnegativity constraint (Hanson, 1988a, 1988b, 1990c). Although the less efficient method E was used in those studies, the principal reason for the ineffectiveness of nonnegativity in the present case is that it is more limited by noise than by the null space. The large amount of noise is needed to limit d_A within the range of reasonable accuracy as discussed in Sec. 3.2. The effects of artifacts were enhanced in previous studies by adding several disks with large amplitude to the scene. In the present study there is only the presence of a second disk with the same amplitude outside a given analysis region. This extra disk can hardly give rise to significant artifacts.

5. Discussion

We have compared several methods for detecting small disks in tomographic reconstructions. The worst performance is provided by method E in which the amplitude obtained by fitting the MAP reconstruction is used as the decision variable. This choice is the same as the matched filter for uncorrelated, Gaussian-distributed, noise fluctuations, so it is more appropriate for unconstrained reconstructions than than for constrained reconstructions. A better decision variable is the mean-square difference between the model and the reconstruction $|\mathbf{f} - \hat{\mathbf{f}}_{MAP}|^2$, as it is similar to a log(likelihood ratio), again ignoring correlations in the reconstruction fluctuations. This method provides much better results. The best detectabilities are achieved by basing decisions on the calculated posterior probability, which takes fully into account the information contained in the measurements as well as in the prior knowledge. In the present tests, however, there is little benefit in re-estimating the exterior region.

We note that as an image containing N pixels, the MAP solution (or a reconstruction of any type) corresponds to a singe point in an N-dimensional space. Any analysis based solely on such a reconstruction must necessarily ignore the complexity of the full posterior-probability distribution, which corresponds to a cloud in the same N-dimensional space. It is the correlations embodied in the posterior-probability distribution that presumably set the ideal observer apart from mortals. A human observer viewing a reconstruction is, in a sense, handicapped by not having access to the full posterior probability distribution and thus may have to resort to the use of a decision method similar to D or E.

The full Bayesian treatment codified by Eq. (7) is expected to represent the ideal observer.

ACKNOWLEDGMENTS. I am indebted to Stephen F. Gull and John Skilling for their inciteful work, especially their MEMSYS 3 Manual (1989), which substantially motivated this work. I have had many stimulating discussions and worthwhile suggestions from Kyle J. Myers and Robert F. Wagner. This work was supported by the United States Department of Energy under contract number W-7405-ENG-36.

REFERENCES

Burgess, A. E. and Ghandeharian, H.: 1984a, 'Visual signal detection. I. Ability to use phase information', *J. Opt. Soc. Amer.* **A1**, 900–905.

Burgess, A. E. and Ghandeharian, H.: 1984a, 'Visual signal detection. II. Signal-location identification', *J. Opt. Soc. Amer.* **A1**, 906–910.

Burgess, A. E.: 1985, 'Visual signal detection. III. On Bayesian use of prior information and cross correlation', *J. Opt. Soc. Amer.* **A2**, 1498–1507.

Butler, J. P., Reeds, J. A., and Dawson, S. V.: 1981, 'Estimating solutions for first kind integral equations with nonnegative constraints and optimal smoothing', *SIAM J. Numer. Anal.* **18**, 381–397.

Gull, S. F.: 1989, 'Developments in maximum entropy data analysis', in *Maximum Entropy Bayesian Methods*, J. Skilling (ed.), Kluwer, Dordrecht, 53–71.

Gull, S. F.: 1989, 'Bayesian inductive inference and maximum entropy', in *Maximum Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, G. J. Erickson and C. R. Smith (ed.), Kluwer, Dordrecht, 53–74.

Gull, S. F. and Skilling, J.: 1989, *Quantified Maximum Entropy - MEMSYS 3 Users' Manual*, Maximum Entropy Data Consultants Ltd., Royston, England.

Hanson, K. M.: 1980, 'On the optimality of the filtered backprojection algorithm', *J. Comput. Assist. Tomogr.* **4**, 361–363.

Hanson, K. M.: 1983, 'Variations in task and the ideal observer', *Proc. SPIE* **419**, 60–67.

Hanson, K. M.: 1987, 'Bayesian and related methods in image reconstruction from incomplete data', in *Image Recovery: Theory and Application*, H. Stark (ed.), Academic, Orlando, 79–125.

Hanson, K. M.: 1988a, 'Method to evaluate image-recovery algorithms based on task performance', *Proc. SPIE* **914**, 336–343.

Hanson, K. M.: 1988b, 'POPART - Performance OPtimized Algebraic Reconstruction Technique', *Proc. SPIE* **1001**, 318–325.

Hanson, K. M.: 1990a, 'Method to evaluate image-recovery algorithms based on task performance', *J. Opt. Soc.* **7A**, 1294–1304.

Hanson, K. M.: 1990b, 'Object detection and amplitude estimation based on maximum a posteriori reconstructions', *Proc. SPIE* **1231**, 164–175.

Hanson, K. M.: 1990c, 'Optimization of the constrained algebraic reconstruction technique for a variety of visual tasks', in *Proc. Information Processing in Medical Imaging*, D. A. Ortendahl and J. Llacer (ed.), Wiley-Liss, New York, 45–57.

Hanson, K. M.: 1991, 'Simultaneous object estimation and image reconstruction in a Bayesian setting', *Proc. SPIE* **1452**, 180–191.

Hanson, K. M. and Myers, K. J.: 1991, 'Rayleigh task performance as a method to evaluate image reconstruction algorithms', in *Maximum Entropy and Bayesian Methods*, W. T. Grandy and L. H. Schick (ed.), Kluwer Academic, Dordrecht, 303–312.

Hanson, K. M. and Wecksung, G. W.: 1983, 'Bayesian approach to limited-angle reconstruction in computed tomography', *J. Opt. Soc. Amer.* **73**, 1501–1509.

Myers, K. J. and Hanson, K. M.: 1990, 'Comparison of the algebraic reconstruction technique with the maximum entropy reconstruction technique for a variety of detection tasks', *Proc. SPIE* **1231**, 176–187.

Myers, K. J. and Hanson, K. M.: 1991, 'Task performance based on the posterior probability of maximum entropy reconstructions obtained with MEMSYS 3', *Proc. SPIE* **1443**, 172–182.

Nashed, M. Z.: 1981, 'Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory ', *IEEE Trans. Antennas Propagat.* **AP-29**, 220–231.

Skilling, J.: 1989, 'Classic maximum entropy', in *Maximum Entropy Bayesian Methods*, J. Skilling (ed.),

Titterington, D. M.: 1985, 'General structure of regularization procedures in image reconstruction', *Astron. Astrophys.* **144**, 381–387.

Van Trees, H. L.: 1968, *Detection, Estimation, and Modulation Theory - Part I*, John Wiley and Sons, New York.

Wagner, R. F., Myers, K. J., Brown, D. G., Tapiovaara, M. J., and Burgess, A. E.: 1989, 'Higher order tasks: human vs. machine performance', *Proc. SPIE* **1090**, 183–194.

Wagner, R. F., Myers, K. J., Brown, D. G., Tapiovaara, M. J., and Burgess, A. E.: 1990, 'Maximum a posteriori detection and figures of merit for detection under uncertainty', *Proc. SPIE* **1231**, 195–204.

Whalen, A. D.: 1971, *Detection of Signals in Noise*, Academic, New York.